International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 7, Issue 2, Apr 2017, 11-16

AN APPROACH TO DETECT TEXT IN VIDEO

MEGHA KHANDELWAL

Assistant Professor, Department of Computer science, University of Delhi, India

ABSTRACT

Detection and extraction of embedded and scene text from video is an important research problem. Its major application is in context based retrieval system. Most of the approaches for this problem make assumption based on the nature of text and are restricted for subclass of videos. The proposed approach detects and extracts text from general purpose video

Received: Jan 08, 2017; **Accepted:** Feb 11, 2017; **Published:** Feb 17, 2017; **Paper Id.:** IJCSEITRAPR20172

Original Article

INTRODUCTION

With the increase in number of videos in multimedia databases the need for an efficient video indexing and retrieval system has now become a necessity. Text present in video carries semantic information about the content in the video and therefore can be used to develop an efficient content based video indexing and retrieval system. Detection of text for developing an indexing and retrieval system is however a challenging task since text in video can be of different size, style and orientation. The change in contrast of background with respect to the text also poses problems during detection of text in video.

The text in video can be classified as scene text and superimposed or artificial text. Scene text is the text which is accidently captured within the scene and hence is an integral part of the scene. Example of scene text include bill boards, number plates etc. on the other hand superimposed or artificial text is the text which is intentionally embedded on the video frame in order to explain the scene in the video frame. The texts in news video, advertisements are some examples of superimposed text.

EXISTING APPROACHES

Features basically used for text detection in video are edge based features [1, 4, 5, 8] and texture based features [9]. Edge based features make use of the rich edge information of text regions for fast text detection, on the other hand, in texture based features the textural features extracted using Gabor filter, wavelet transform, gradient orientations are used for detecting text in video frames.

The proposed approaches for video text detection by the researchers can be broadly classified as heuristic, machine learning and hybrid. Heuristic approach use empirical rules and thresholds based on edge features in order to distinguish between text and non text areas. Shivakumara *et al.* [1] proposed a Heuristic method based on filters and edge analysis for identification of block containing text. Min Cai *et al.* [4] also proposed a heuristic approach using features such as edge strength, edge density and horizontal distribution.

Machine learning approach are based on classification techniques trained on text and non text patterns which scan the video frame in order to localize the occurrence of text. Xiaojun Li *et al.* proposed an approach where 24 features extracted using stroke filters in horizontal, vertical, left diagonal and right diagonal directions

are extracted and SVM classifier is used to verify the candidate text lines. Wenicke *et al.* also proposed an approach where a neural feed forward network which was trained using 7 training cycles is used for developing classifier.

Hybrid approach usually consists of two stages where initial text localization is done using fast heuristic method which is followed by verification of previous results and eliminating the false alarms. Anthimipoulos *et al.* proposed a hybrid approach consisting of two stages. In the first stage morphological operations are performed on canny edge map of video frame, followed by bounding box generation and splitting of text area to text lines using projection profile. In the second stage results are verified using edge local binary pattern (eLBP). Ye *et al.* [8] proposed an edge based method followed by a texture based method for text detection in video frames.

Heuristic approach is fast as compared to other approaches but there are some geometrical constraints involved based on characteristics of text while using this approach. On the other hand machine learning approach is computationally complex. Hybrid approach uses heuristic approach along with machine learning approach for better results.

PROPOSED APPROACH

The input to our system is a sequence of video frames captured from the video. In our system morphological operations are performed on edge map of video frame generated using canny edge detection algorithm, which is followed by initial bounding box generation using contours. Bounding box are then refined based on their area and dimensions. After which, text area is decomposed to text lines using horizontal and vertical projection profiles. Lastly, verification of bounding box is done using skeletonization and horizontal alignment of text. Basic steps in our system are shown in Figure 1.

Canny Edge Detection

Our algorithm uses the fact that text line produce number of vertical edges which are horizontally aligned. Initially, the input video frame is converted to grayscale for further processing. Then, binary edge map of the obtained grayscale frame is generated using canny edge detection algorithm.

Morphological Operations

In this step, morphological dilation followed by opening operation is performed on the edge map of input frame in order to connect vertical edges and remove false alarms.

Initial Bounding Box Formation

This step consists of bounding box formation for every non-zero valued connected component, which are known as initial candidate box.

Bounding Box Refinement

The bounding box obtained in the previous step is removed if:

- The bounding box is on the border of the frame.
- The bounding box is inside another bounding box.

This step removes the bounding box formed because of the frame border and also the small bounding boxes inside some other bounding box.

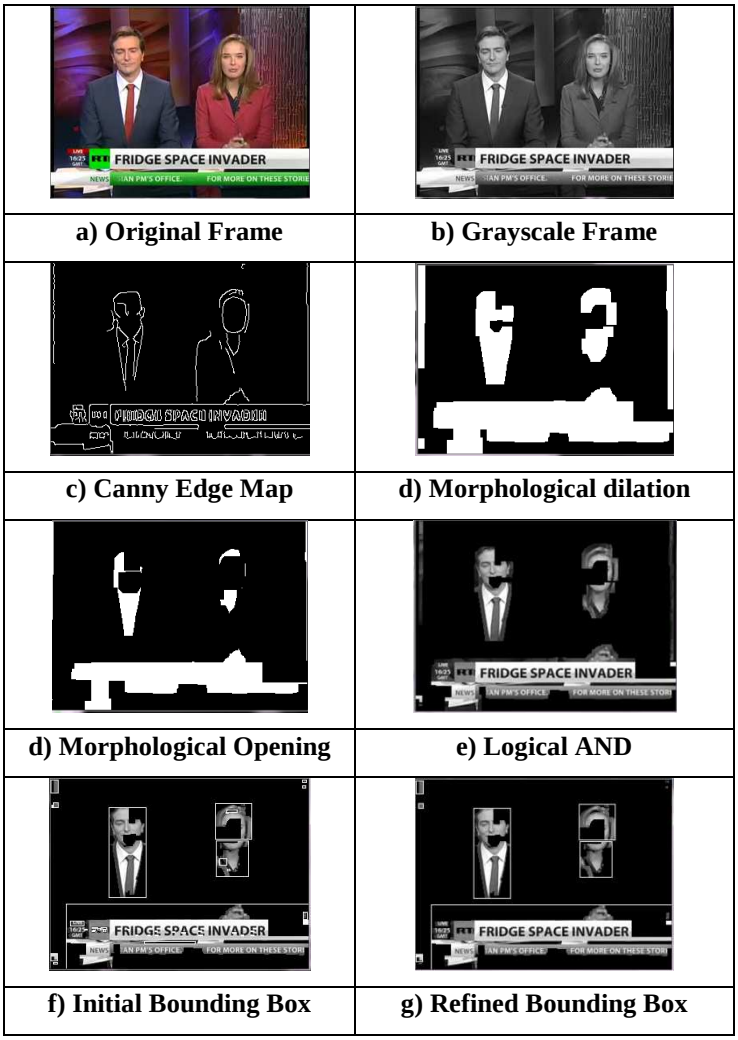

Figure 1: Steps of Proposed Approach

Region Decomposition

Horizontal/ Vertical projection profile is the sum of intensities of pixels over column/ row which can be used for segmentation of text in the document.

The bounding box obtained in previous step consists of text area. This text area is decomposed into text lines and words by thresholding horizontal and vertical projection profiles of the bounding box. Here, we used adaptive thresholding in order to reduce false alarms.

Figure 2 (b) shows projection profile of Figure. 2(a) which is depicted as histogram along y axis. If profile graph of a bounding box is lower than certain threshold at some location, than that bounding box is divided into two boxes from that location as shown in Figure 2(c). Same is done with vertical projection profile.

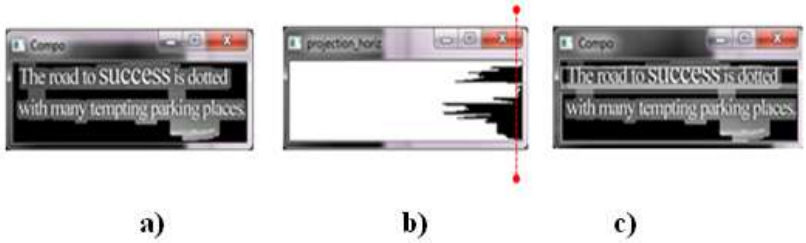

Figure 2: Region Decomposition

Bounding Box Verification

In our system bounding box verification is done in two steps.

In first step morphological operations are performed on edge map of bounding boxes obtained after region decomposition. Then, the skeleton of the processed box is created and the number of white pixels is calculated. If this count is greater than a certain threshold than the bounding box is removed since it is considered to be a false alarm.

In second step the bounding box are verified based on the fact that text lines consist of characters which are horizontally aligned. If there are more than two bounding box with same y coordinates than these bounding box are considered of having text, forming a text line. Otherwise edge map of bounding box is vertically decomposed and if a bounding box is found to have more than 3 bounding box than it is considered of having text, forming a word. Bounding box which do fulfill above mentioned criteria are considered as false alarms and are therefore removed.

RESULTS

Text blocks detected using text detection method can be categorizes as:

- Truly Detected Block (TDB): A detected block that contains a text line, partially or fully.
- False Detected Block (FDB): A detected block that does not contain text.
- For each frame in a video we manually count the Actual Text Blocks(ATB)

The performance measures are defined as follows:

- Detection Rate (DR) = TDB /ATB
- False Positive Rate (FPR) = FDB / (TDB+ FDB)

The proposed approach was tested on 10 different videos, where duration of videos varied from 1 min to 5 min. Frame per second of videos was 25 to 29 fps.

Overall detection rate was measured as 95.4% whereas false detection rate came out to be 7.23%.

CONCLUSIONS

The proposed approach is able to detect text within the range of 5 pixels to 25 pixels. Efficiency is best when video contain high contrast text with simple background is detected. The proposed approach works for only horizontally aligned text

REFERENCES

1. *Palaiahnakote Shivakumara, Trung Quy Phan and Chew Lim Tan "Video Text detection based on filters and edge features", IEEE International Conference on Multimedia and Expo, pp 514-517, 2002.*
2. *Trung Quy Phan, Palaiahnakote Shivakumara and Chew Lim Tan "A laplacian method for video text detection" 10th IEEE International Conference on Document analysis and recognition, 2009.*
3. *Johann Poignant, Franck Thollard, Georges Quenot and Laurent Besacier, "Text detection and recognition for person identification in videos", 9th IEEE International Workshop on Content-Based Multimedia Indexing (CBMI), pp 245-248, 2011*
4. *Min Cai, Jiqiang Song and Michael R. Lyu, " A new approach for video text detection", IEEE International Conference on Image Processing, 2002.*
5. *Jie Xi, Xian-Sheng Hua, Xiang-Rong Chen, Liuwenyin, Hong-Jiang Zhang "A video text detection and recognition system", IEEE International Conference on Multimedia and Expo, pp 876-876, 2001*
6. *Axel Wernicke and Rainer Lienhart, "On the segmentation of text in videos", IEEE International Conference on Multimedia and Expo Vol: 3, pp 1511-1514, 2000*
7. *Marios Anthimopoulos, Basilis Gatos, Ioannis Pratikakis "A hybrid system for text detection in video frames", IEEE, The Eighth IAPR International Workshop on Document Analysis Systems, pp 286-292, 2008*
8. *Qixiang Ye, Wen Gao1, Weiqiang Wang, Wei Zeng "A robust text detection algorithm in images and video frames", Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing and Fourth Pacific Rim Conference on Multimedia,Vol. 2, pp 802-806, 2003*
9. *Palaiahnakote Shivakumara, Trung Quy Phan, Chew Lim Tan, "New wavelet and color features for text detection in video", 20th IEEE International Conference on Pattern Recognition(ICPR), pp 3996-3999, 2010.*
10. *Xian-Sheng Hua, Pei Yin, Hong- jian Zhang, "Efficient video text recognition using multiple frame integration", IEEE International Conference on Image Processing, Vol. 2 2002.*